Conclusion

11.1 Introduction

The current chapter reviews the various phases of this work, and summarises the results of the analyses. Furthermore, it discusses the limitations of the current approach, and suggests possible extensions and ideas for future work. Despite its several limitations, the work may be considered to contribute to the current state of knowledge and research in corpus linguistics and cultural studies in several ways which will be illustrated in a concluding section.

The following paragraphs review the aims, theoretical background, and research questions that guided the investigation.

In a general attempt to contribute to our understanding of cultural systems, and of the relationship between text, semantics, and culture, I selected and outlined two models of culture which lend themselves to semantic as well as quantitative analyses; these are the systemic models by Fleischer (1998) and Nobis (1998), described in Chapter 2. In particular, according to Fleischer, discourse, i.e. the linguistic level at which culture shows itself and develops, is characterised by symbols. In turn, symbols are composed of three elements: the core, which is a stable semantic element that is shared by all members of the cultural community; the current field, a semantic element which is shared by several, though not all, members of the community but which is spreading; and the connotational field, a semantic element that is specific to single individuals. Both core and current fields are expressions of cultural meanings and can be identified by analysing the frequency and distribution of the semantic associations of a given concept/word across the members of the cultural community. Consequently, Fleisher's theory will help us establish the level at which a concept (expressed by a word) is rooted (or anchored in Fleisher's terminology) in a given culture at a given point in time. Nobis' theory on the other hand will help us compare two cultures in terms of their relative development with reference to the same symbol. In fact, Nobis banks on the generalized systemic idea that systems are in constant tension between stability and evolution – the latter being achieved by transmission of behaviours (including mental behaviours) – and suggests that transmission of behaviour may only take place when that behaviour has a long established network of relations with other behaviours, i.e. a stable behavioural pattern. Nobis' notion of 'stable behavioural pattern' can easily be equated with Fleischer's notion of conventionalisation: a mental behaviour, such as thinking of a concept, has features

that, at a given point in time, are widely shared by all members in a community/culture. Consequently, if two cultures show relevant differences in the number of the concepts connected to the same key concept we can hypothesise different stages of knowledge/acceptance of the key concept. These theories were tested on two concepts, *chocolate* and *wine*, in the English and Italian cultures. These concepts were selected, among other considerations, because one of them (*wine*) could be expected to have different rooting in the two cultures, while the other (*chocolate*) to have similar rooting, given the two countries' climatic conditions and food production histories.

Furthermore, a review of semantic approaches to the study of culture in different disciplinary areas, including linguistics, anthropology, and psychology (Chapter 2), suggested elicited data and non-elicited data as equally possible materials for cultural analysis. The same review also suggested that corpora and quantitative analytical methods are easily applicable to this purpose. In particular, the most frequently used analytical methods seem to revolve around the use of frequency lists, keyword and collocate frequencies, as well as grouping of items to create superordinate domains (either entirely semantic or thematic). These topics, along with other major issues in corpus linguistics such as the use of the Web as a source for corpus data, were detailed and discussed Chapter 3.

Finally, given my desire to perform an analysis of cultural associations of a given concept which could find theoretical or practical applications not only in the linguistic and cultural fields, but also in the marketing one, Chapter 4 overviewed the materials and methods most frequently used in marketing research, reviewed selected marketing and consumer studies where analysis of linguistic data is performed, and established some methodological common ground among linguistics, cultural studies, and marketing. Such common ground can be summarised in the following features: use of elicited data, but also of Web data; analysis of word associations; semantic/content analysis; and frequency as a measure of the association's importance.

The theoretical and methodological elements outlined above provided the framework for the experimental part of the work. This part of the work – which focussed on the development of a suitable analytical method to establish and compare the cultural mental associations of *chocolate* and *wine* in Great Britain and Italy, and on testing different types of datasets for cultural analysis – was operationalised in five Research Questions. Research Questions 1 and 2 – What are the semantic associations of chocolate, and wine in the Italian and English cultures?; and What are the differences between the Italian and English cultures with reference to chocolate, and wine?, respectively – were addressed in Chapter 6, and the results of that Chapter are summarised in Section 11.2.1 below. Research Question 3 – Could we identify the cultural associations of the two words without coding the entire dataset? – was addressed in Chapter 7 (on elicited data) and Chapter 8 (on Web data); the results of the analyses performed are summarized and discussed in Section 11.2.3, below. Research Question 4 – Could we identify the cultural associations of the two words using an automatic semantic tagger? – was addressed in Chapter 9 and is summarised in Section 11.2.4, below. Finally, Research Question 5 – Could we identify the cultural associations of the two words using a general (Web) corpus? – was addressed

in Chapter 10, and the results are discussed in Section 11.2.2, below. Please notice that in the current chapter the sections have been organized in a way that is slightly different from the progression of the research questions. This was done because, at this stage, it seemed useful to highlight the cultural part of the work separately from the purely methodological one.

11.2 Summary of the experimental work and results

The current experimental work can be divided into two logical parts. The first part is concerned with finding the most suitable materials and methods for the analysis of linguistic data to highlight cultural features. This has been subdivided here into two separate sections, summarising the results of elicited data analysis (Section 11.2.1) and of Web data analysis (Section 11.2.2), respectively. The second logical part is concerned with methodological issues, such as testing different sampling procedures in order to avoid having to code large datasets (Section 11.2.3), and using an automatic semantic tagger in place of manual coding of the data (Section 11.2.4).

11.2.1 Retrieving cultural associations in elicited data (control situation)

The current work used elicited data on *chocolate* and *wine*, gathered through free sentence-completion and sentence-writing tests in English and Italian, to highlight the cultural associations that each key word has in the cultures considered. The elicited data were manually analysed using content analysis procedures (i.e. semantic coding), and the semantic categories that emerged from the content analysis were quantitatively measured in terms of overall frequency, as well as frequency distribution across subjects – the latter being calculated by applying Molinari's evenness index. Furthermore, by looking at the position of the evenness index with reference to the confidence interval, it was possible to establish the level of conventionalisation of the various fields and domains, in each culture and for each node word. Three conventionalisation levels were considered: high, medium, and low, respectively corresponding to Fleischer's core, current and connotational fields. Finally, the results of the Italian and English datasets were compared by the Welch *t* Test for Independent Samples.

In keeping with expectations, *chocolate* appeared as an equally long- and well-established symbol in the two cultures; on the other hand, *wine* – though well-established in both countries – showed longer rooting in the Italian culture, as the Italian respondents' answers showed a remarkably higher percentage of highly conventionalised semantic fields and domains, and a remarkably lower percentage of low conventionalisation fields and domains than the British ones.

In the light of the current experiments on elicited data, the Italians seem to distinguish themselves from the British for their more frequent matching of *chocolate* to the following concepts: BAKERY/COOKING; RECIPE; DIETING; MEDICINE; BEAUTY; HISTORY; NICE/PLEASANT/PLEASURE; CHILDREN; FAMILY; STUDYING/INTELLECT; QUALITY/TYPE; GENUINE. On the other hand, more prominent for the English than for Italians seem to be: WOMEN; and PRICE.

As regards *wine*, the following semantic fields emerged as distinctively more prominent for the Italians than for the English: BAKERY/COOKING; EVENT; WOMEN;

NATURE; ARTISTIC PRODUCTION; QUALITY/TYPE; QUANTITY; GENUINE; PRICE. On the other hand, more prominent for the English than for the Italians were: PRODUCT/SHAPE; DRINK; MANUFACTURING; RECIPE; LANGUAGE; CONFIDENCE; DESIRE; NICE/PLEASANT/PLEASURE; MEN; FRIENDSHIP; POSH; SHARING/SOCIETY; PEOPLE; and STUDYING/INTELLECT.

As we have argued in Chapter 6, this is only a list of the mental associations in which the English culture seems to differ from the Italian one. Neither the qualitative nor the quantitative analyses performed in this work can in any way explain the type of the association or the reasons for the differences. Further steps, such as analysis of individual concordance lines, are needed to understand the exact link between key word and semantic field in each culture. Such analyses are beyond the scope of the current investigation, but will be considered in future extensions of this work. Nevertheless, I believe that analyses of this type may be adopted in the exploratory phases of marketing (or cultural) research, where research aims to outline problems, collect information, eliminate impractical ideas, and formulate hypotheses.

11.2.2 Comparing Web data to the control situation

The elicited data – considered as the control situation – were compared to non-elicited sentences on *chocolate* and *wine* from general Web corpora in English and Italian.

The Web corpora – analysed through randomly sampled subsets of about 1800-2000 sentences – retrieved over 90% of the semantic fields with high conventionalisation and of the cultural associations attested in the corresponding elicited datasets, and 100% of the domains. However, the corpora also retrieved most of the low conventionalisation fields, along with a few extra fields whose conventionalisation level is not known (although one could speculate that – being those fields totally absent in the elicited corpora – they could be classified as having low conventionalisation). The Web results were quantitatively compared to the elicited ones by means of Spearman's Rank Correlation Coefficient and showed modest correlation at the level of semantic fields, and strong correlation, or higher, at the level of conceptual domains. Furthermore, no marked and systematic differences can be seen between the results of the English data vs. those of the Italian data.

Finally, the ASSESSMENT field matched, in ranking, the results of the elicited datasets, with positive assessment preceding neutral assessment, which in turn preceded negative as well and undecided assessment results. Interestingly, however, the Web sub-corpora systematically showed percentages of negative assessment which are remarkably lower than those in the elicited datasets, a result which is at least partly connected to the 'marketing flavour' of large part of the texts in the Web corpora – the latter being also a probable explanation for about 30% of the semantic fields present in the Web corpora, but absent in the corresponding elicited datasets.

Consequently, despite our initial fears that issues such as uncontrolled authorship and readership (see Chapter 3) could represent a bias in the use of English Web data, comparisons between the Web corpora and the elicited data suggest that large general Web corpora can be considered representative of the cultural associations of a node word. In fact, randomly sampled Web subsets of only 1800-2000 sentences, included all the relevant cultural associations of the node word.

Furthermore, when the coding scheme adopted was broad and included few categories, the general Web corpora appeared to be representative not only at a qualitative level, but also at a quantitative one.

Unfortunately, as argued in Chapter 6, we cannot rely on frequency alone to establish conventionalisation. Only the very highest ranks in the frequency list are systematically occupied by low conventionalisation fields, and only the very lowest ranks are systematically occupied by high conventionalisation ones. Any other position in the list can hardly tell us something about conventionalisation level. Consequently, if we had only Web data, and no control elicited data, we would have to assess the conventionalisation level of each field/domain by applying an evenness index, in order to establish which of the retrieved semantic fields/conceptual domains can be safely considered cultural associations. Fundamental pre-requisite for applying the evenness computation is the possibility to group the Web sentences according to subject/author or website. This – along with T-test analyses for cross-cultural comparisons – could not be done in the current work, because at the time when the Web data were retrieved, the Sketch Engine did not provide information about the website each text was taken from. The updated version of Sketch Engine, however, does provide this type of information, and its users can now benefit from the possibility to assess the distribution of concordance lines across Web sites (i.e. authors).

11.2.3 Testing different procedural approaches

The current work experimented different procedural approaches. In particular, focus was on finding an alternative route to manual coding of the whole dataset, as this is a costly and complex procedure when the number of sentences in the dataset is very high. The various procedures were all tested on the elicited datasets, while only the procedures that had showed better results were applied to the Web datasets .

One of the procedures adopted was random sampling of the sentences in the dataset, a rather standard procedure to create smaller, but representative sub-sets. Kilgarriff (2001b) suggests generating several random samples and average the results, to guarantee maximal representativeness of the sample; in the current work multiple random sampling will be substituted with sampling on different data sets followed by assessment of the consistency of the results.

The other two procedures were based on analysis of a limited number of the most frequent words in the datasets. These less standard procedures were inspired by previous linguistic studies of culture and by Fleischer's theories which suggest the existence of a relationship between cultural associations, their level of conventionalisation and frequency of occurrence of the given associations. This led me to testing the following two possibilities: 1. performing manual semantic analysis of the most frequent 50/100/150/200/250/300 content words in the wordlist, by generating concordances for each word, reading through the concordance lines and matching each word to one or more of the semantic categories available; and 2. using the four most frequent content words to extract sentences from the manually coded dataset and create a sampled sub-corpus.

The random sampling technique proved to be the most representative route, as it systematically showed higher results that the others at all levels of analysis,

including separate analysis of semantic field ASSESSMENT. In fact, the randomly sampled corpora, identical in size to the 4-lemma ones, showed 79-94% of the semantic fields in the datasets, corresponding to 96-100% of the highly conventionalised fields and 94-98% of the cultural associations, with correlation results falling in the very strong range. Furthermore, this procedure showed qualitative and quantitative results that were perfectly comparable to those of the whole dataset as regards analysis of the ASSESSMENT category.

The other two techniques provided interesting results at both qualitative and quantitative level, but not when it came to analysing semantic field ASSESSMENT. In fact, the top 300 content words retrieved 65-70% of the total number of semantic fields in the whole datasets, 86-94% of the highly conventionalised fields and an almost identical percentage of the cultural associations, with correlation results in the strong range. The top four words in the frequency wordlist, treated as lemmas, provided sub-corpora whose size varied between 25% and 35% of the corresponding original dataset and showed 72.6-83% of the semantic fields in the datasets, corresponding to over 95% of the highly conventionalised fields in the original datasets, and 94-96% of the cultural associations, with correlation results in the strong-very strong range. Separate analysis of the ASSESSMENT category, however, showed qualitative and quantitative results that were not comparable to those of the whole dataset.

11.2.4 Testing the use of an automatic semantic tagger

Finally, an automatic semantic tagger (Wmatrix/USAS tagset) was tested on the elicited data, in order to assess the extent of its possible application in cultural analysis. The automatic semantic tagger was used in two different scenarios: 1. an 'autonomous' scenario, where the USAS tagset was automatically applied to the elicited and Web datasets and the results of the tagging process were compared to the most frequent content words in their wordlists, and to sub-corpora randomly sampled from the same datasets; and 2. a 'comparative' scenario, where the USAS tags retrieved in the datasets were converted into Codebook tags and results were compared to those of manual coding.

In the 'autonomous' scenario, it was noticed that by applying USAS tagging the most frequent content words in the wordlists and the sub-corpora randomly sampled from the datasets both retrieved a smaller percentage of semantic fields than with manual coding. This is a consequence of the very high granularity of the USAS dataset which – with all its categories and subcategories, as well as 'pluses' or 'minuses' to indicate a positive or negative position on a semantic scale – includes almost 400 different labels. Interestingly, however, although the percentage of semantic fields retrieved was lower when USAS tagging was applied (about 44.5% vs. 68-70% when analysing the most frequent words in the wordlist; 66.79% vs. 84.09% for *chocolate*, and 80.36% vs. 86.9% for *wine*, when analysing the random sub-corpus), in both cases Spearman's test results were very similar to those obtained with manual tagging, the correlation index being always in the strong range. Finally, when the USAS tagset was applied to the Web corpora, it became evident that the number of most frequent semantic words in the wordlist necessary to highlight the most frequent semantic associations of the node word depends on corpus size. In fact, the top 300 word retrieved only about 25% of

the semantic fields. Even extending to 450 the number of words considered, the percentage of fields retrieved was still very low (about 31%). However, correlation values were in the medium range and showed constant linear increase, thus strengthening the hypothesis that the most frequent words in corpus are highly representative of the contents of the corpus.

In the 'comparative' scenario, the English elicited datasets were automatically tagged with Wmatrix, and the most frequent 150 items in the resulting semantic frequency lists were compared to: A. the results of manual coding of the entire datasets, and B. manual coding of the most frequent 300 words in the wordlist. In both cases, a conversion scheme which matched the USAS tags to the semantic fields used in the manual coding of the elicited data was applied. Comparison showed that the most frequent 150 items in the USAS frequency list – which represent 56% of each list – showed about 67-68% of the Codebook fields highlighted with manual tagging, and about 93% of the conceptual domains, including 74-80% of the highly conventionalised fields and about 80% of the cultural associations, and 100% of the highly conventionalised and cultural domains. Furthermore, the most frequent 150 USAS categories in the semantic frequency list showed marked preference for positive, rather than negative assessment, as was the case in the control situation. From a quantitative perspective, correlation results assessed using Spearman's test showed modest correlation for semantic fields and modest/strong correlation for conceptual domains. This is most certainly due to the quantitative approximations adopted in the conversion procedure. In fact, in about 34% and 30% of the cases, for semantic fields and conceptual domains, respectively, the frequency of the USAS tags considered was equally (and not proportionally) distributed among two or more Codebook semantic fields, which obviously influenced Spearman's results. Finally, the most frequent 150 USAS items in the semantic frequency list proved to be less representative of the whole dataset than manual coding of the most frequent 300 words in the wordlist.

11.3 Some methodological considerations

Methodological issues were a major concern in the current work from the very begging. A summary of the methodological considerations emerging from the investigation is provided in the following paragraphs.

The current research confirms that, alongside elicited data, which are a typical source of linguistic material in marketing research, the Web can be a useful source of data for analysing cultural associations of a given word or concept. The current work tested freely available large general Web corpora, from which sentences containing the word under analysis were extracted. In the current research, it was not possible to compute evenness measures in the Web corpus, and the results could only be interpreted by comparing them to the elicited data. The comparison, however, showed that a small random sample of the Web data included all the relevant cultural associations of the node word. This leads to believe that if the Web data are collected in a way which allows the researcher to group the Web sentences according to subject/author or website, the Web data could be interpreted regardless of the presence of 'control' data.

Furthermore, the research suggests that a relatively small number of sentences including the given key word is sufficient to understand its cultural associations. In fact, the current research tested eight random sub-sets sized 20-30% of the original datasets, and including from about 400 to about 2000 sentences. Each of them retrieved about or over 95% of the cultural associations. This procedural finding is particularly relevant when dealing with a very large dataset including several thousand sentences. Indeed, manual coding is not only time-consuming, but also highly demanding: it requires the work of at least two well-trained coders, as well as an intense effort from each of them in terms of coherent and cohesive application of the given coding scheme. And the larger the dataset, the greater the effort in applying the scheme consistently and coherently. Also an analysis of the most frequent words in the wordlist could, if necessary, be employed as an alternative route to tagging the whole corpus, bearing in mind – however – that the number of words to consider depends on the size of the original corpus and that this procedure introduces approximations. The effectiveness of these two procedures are easily explained by the fact that cultural associations emerge from a combination of frequency and spreading across a large number of subjects.

Finally, if a suitable automatic semantic tagger is available, quick and consistent semantic analysis of the whole corpus can be easily obtained, and the cultural associations can be identified by looking at the most frequent semantic categories in the corpus. However, if the corpus under analysis is small, such as the elicited ones used in the current work, the use of an automatic semantic tagging tool is recommended only if the semantic categories of the automatic tagging can be used without further conversion, since conversions introduce approximations.

As regards approximations, however, it must be said that, as Hubbard (2010, p. 23) clarifies, measurement is "a quantitatively expressed reduction of uncertainty based on one or more observations" and, in many circumstances, having even an approximate idea of the variables and their values represents a big leap ahead from our original level of knowledge about the given object. This is indeed the case of exploratory market research, where – as we have seen – the researcher's aim is to acquire an inexpensive approximation and uses it to outline problems, eliminate impractical ideas, and formulate hypotheses.

A further consideration regards tagset granularity. In the current work, three different tagsets were applied to the same data: the Codebook semantic field tagset, including 96 semantic fields; the Codebook conceptual domain tagset, with its 16 conceptual domains; and the USAS tagset which includes almost 400 different tags. The Web data, as well as the elicited wordlists and sampled sub-corpora were compared to the 'control' data after applying each of the three tagsets. Throughout the work it consistently appeared that when passing from a more detailed to a less detailed tagset (e.g. semantic fields vs. conceptual domains; USAS tagging vs. manual tagging), semantic category coverage increased, and also correlation values increased. This is in keeping with observations by Guerrero, Claret, Verbeke *et al.* (2010, reviewed in Chapter 4), who applied to their data a double grouping process with categories which are comparable to our semantic fields and conceptual domains and noticed that greater differences between cultures appeared at the level of semantic fields. Furthermore, I also share their considerations about the advantages and limits

of grouping semantic categories into conceptual domains, when they say that its main advantage

> "is its simplicity, although the double grouping process increases the subjectivity of the results obtained. In addition, some difficulties may be observed when trying to obtain a reduced number of classes because it was not always easy to group the different classes together under a common dimension or concept. It is also important to notice that using this approach the more subtle differences between regions may disappear" (*ibid*., p: 230).

I would add, however, that the grouping of semantic fields into conceptual domains facilitated choosing one semantic field over another when performing manual tagging. So I would suggest creating and using a two level tagging scheme when coding the data, but limiting analyses to the more fine-grained level in the tagset.

Finally, a look at the semantic fields which are absent with reference to both key words in the same culture suggests that field presence/absence may depend on the key word, rather than the culture. In fact, only one field is systematically absent in the English datasets (COMPETITIVENESS), and none in the Italian ones. It must be remembered that, in the current work, the overall numer and range of fields and domains emerged from the data themselves. Consequently, absence is relative to the coding scheme; any semantic field which does not appear in the Codebook could be considered 'absent' in both cultures and for both node words. Nevertheless, I would tentatively declare that this finding supports the use of dedicated coding systems for different node words. However, the relationship that links presence/absence of a semantic field and culture requires further investigation, on a much wider number of node words.

11.4 Limitations of current work and future directions

The current work has some limitations and possible directions for development have already been identified.

A major limitation derives from not having controlled the composition of the two population samples when collecting the elicited data. Although the English and Italian groups of respondents show some overall similarities (a majority of university students in the 18-25 age range; data collected in both Northern and Southern areas of the two countries), no precise data was available in the current research as regards variables such as the respondents' age, gender or occupation. The fact that the elicited data and the Web data analysed in Chapter 10 showed similar results, to some extent confirms similarity between the two population samples. Further confirmation could be found by applying, one or more of the following:

1. Replication of the study, possibly also with a larger sample size and/or more stratified random sampling.
2. Other elicitation methods (e.g. story writing).
3. Depth interviews and focus groups, possibly with deliberate attempts to elicit and probe the concepts that showed cultural differences (e.g. ask Italian and English respondents deliberately about women and chocolate and see if there is a difference in how they talk about the subject).

4. Content analysis (visual as well as verbal) of representative samples of chocolate/wine advertising from UK and Italian companies addressing the local audience.

For the time being, we will have to accept these results as they are. Should further research disconfirm this cultural comparison and cast doubts on the frequency-plus-T-test method adopted here, nevertheless, the methodological investigations performed in comparing different types of data and/or coding schemes will still be valuable.

Next, the analyses performed in this work highlight the semantic categories which are culturally connected to the given key words, but do not allow the researcher to understand the kind of relation that exists between the category and the key word. For example, when Italians talk about *children* and *wine*, what exactly do they refer to: that wine can or cannot be given to children? A possible way to answer this question would be to look at concordance lines. An analysis of the concordance lines of each semantic category could represent an interesting extension to the current work, and might provide greater insight into cultural specificities.

Third, the analysis of the ASSESSMENT field performed in this work, although clearly limited in scope, was sufficient for the purposes of the current work and was a suitable reference term for the methodological comparisons which were performed in the various chapters. However, from the perspective of cultural and even more so consumer research, the current level of analysis of semantic prosody appears excessively limited. We may expect different cultures to focus on different features when positively or negatively assessing a concept. Furthemore, we noticed in Chapter 4 how sentiment analysis plays a fundamental role in marketing research. Opinionated text may, for example, orient consumer behaviour when purchasing products, warn marketing managers about the rising of critical situations, or help establish the pricing power of a product feature. Consequently, in order for assessment analysis to find any application in marketing, even in exploratory phases, it needs to investigate further factors such as the reasons behind positive/negative evaluation, and the features of the concept/product which triggered the evaluation. With reference to the procedures and tools adopted in the current work, an extension of the analysis of the ASSESSMENT field could see the application of the following analytical methods: 1. looking at the distribution of the Positive and Negative categories across the various fields/domains;[1] 2. analysing the evaluative adjectives that collocate with the two selected key words;[2] 3. retrieving key relations between words, such as "the attributes assigned to various persons or things, and the various modifying and negating words and phrases associated with these" (Wilson, 1993, p. 6).[3]

Fourth, in the current research, it was not possible to compute evenness measures in the Web corpus, and the results could only be interpreted by comparing them to the elicited data. The comparison, however, showed that a small random

---

[1] A quick look at the data suggests that, when performing this type of analysis, it will be important to consider only the semantic fields/domains which show a minimum number of hits, alongside a significant a difference between Positive/Negative Assessment.

[2] Methodological inspiration could be taken from the works by Baker (2006), reviewed in Chapter 2, and Aggarwal, Vaidyanathan and Venkatesh (2009), reviewed in Chapter 4.

[3] In previous versions of the Wmatrix system, features for the automatic retrieval of key relations were available (Wilson 1993). Unfortunately, in the on-line version these features are no longer available.

sample of the Web data included all the relevant cultural associations of the node word. This leads to believe that if the Web data are collected in a way which allows the researcher to group the Web sentences according to subject/author or website, the Web data could be interpreted regardless of the presence of 'control' data.

Furthermore, the current analyses concerned two consumables. It might be interesting to test the procedures and automatic semantic tools described in this work on other topics of cultural and/or marketing interest. A relevant candidate is certainly node word *traditional* in food talk, since the results could be compared to those of Guerrero, Claret, Verbeke *et al.* (2010) in their study on the perception of traditional food products (see Chapter 4). Another node word that comes to mind is *flexibility*, as this word seems to have different semantic prosodies in the Italian and Anglo-American cultures: Italians seem to praise 'flexible procedures' (almost an oxymoron for the English or Americans), and strongly oppose 'flexibility' when it refers to the need to adapt themselves to a constantly changing job market.

Finally, in the current work experimentation with automatic tagging was possible only for English, since no semantic tagger based on a coding scheme similar to that of Wmatrix exists for Italian. Currently, a Finnish and a Russian version of the USAS tagset exist, alongside the English one.[4] It would be interesting to develop an Italian USAS tagset and test it on the *chocolate* and *wine* data.

11.5 Contribution to knowledge and concluding remarks

Despite its clear limitations, the current work can be considered to contribute to knowledge in several ways.

First of all, the current work is characterized by an interdisciplinary perspective which links linguistics, marketing research and cultural studies. The combination of the three fields seems innovative and certainly provides interesting methodological as well as theoretical ideas from which all the three disciplines could benefit.

Second, the quantitative comparisons between the entire datasets (elicited as well as Web) and smaller samples of the data accomplished in this work add useful pieces of information to our general knowledge in corpus linguistics.

Third, the procedure adopted to establish the cultural associations of the key words was specifically developed after careful analysis of similar experiments described in the scientific literature of different disciplines (linguistics, cultural studies, and consumer research) and in the light of the cultural systems theories by Fleischer (1998) and Nobis (1998). In particular, Fleischer (1998) suggests a quantitative type of analysis based on frequency and spreading of specific individual phenomena. Consequently, I believe that the procedure adopted in the current work represents an improvement to previous frequency-based measurements of cultural semantic associations. In fact, mere observation of raw or mean frequency of fields and conceptual domains provides an approximate picture of the semantic associations, as it disregards distribution of answers across subjects. On the other hand, the use Molinari's evenness index – inspired by Wilson and Mudraya (2006), but here applied in a different way – introduces a quantification of spreading. Confirmation of the

---

[4] See http://ucrel.lancs.ac.uk/usas/

validity of the procedure applied comes from the fact that the results are in keeping with expectations.

Finally, this work is among the rare applications of Fleisher's theory of culture, and provides results which seem to support the theory. Furthermore, Nobis' system theory is also confirmed. In fact, the *wine* experiment clearly confirmed that longer standing of a concept (*wine*) in a given culture (Italy) corresponds to stronger cultural rooting (Nobis, 1998), here expressed in terms on higher percentage of highly conventionalised semantic fields. The second of Nobis' hypotheses, postulating greater semantic complexity of longer standing concepts, is supported in the *wine* experiment not by the overall number of semantic fields associated to the given concept, but by the greater number of semantic elements which are shared by several respondents, i.e. those semantic fields or conceptual domains with high level of conventionalisation. These two system theories, though still little known among linguists and consumer researchers, have much to offer to cultural analysis. Furthermore, they lend themselves to quantitative research and, thus, to corpus linguistics.

To conclude, I believe that the current work has been rather successful in its aim to contribute to our understanding of cultural systems, and of the relationship between text, semantics, and culture. Furthermore, it provides theoretical as well as practical ideas for improving cultural analysis through language.

All the three main areas of studies considered in this interdisciplinary research may benefit from its theoretical reviews and discussions and the results of its analyses. In particular, I believe that analyses of the types performed in the current work could be adopted in the exploratory phases of marketing or cultural research, where research aims to outline problems, collect information, eliminate impractical ideas, and formulate hypotheses.